Hierachical and Integrated Algorithms: Comparison and Applications in Motion Estimation and Recognition[1]

Kun Huang and P.R. Kumar
Dept. of Electrical and Computer Engineering, and Coordinated Science Laboratory
University of Illinois, 1308 West Main Street Urbana, IL 61801, USA
{kunh,prkumar}@decision.csl.uiuc.edu

Abstract

In this paper, we study several issues in motion estimation and object recognition. First, we compare the performance of two hierarchical and integrated methods in motion estimation. Second, we address the use of a simulated annealing algorithm for object recognition. This algorithm is then adapted for vehicle identification.

1 Introduction

The structure of a information processing system can be classified as a hierarchical structure, an integrated structure, or a mixed one.

In a hierarchical or layered system, data processing is divided into different levels. In each level, some abstract information is extracted and sent upwards as the input to the next/higher level. Hence, the higher the level, the closer we are to a conceptual understanding. A vision system is a good example of a hierachical system [1]. One advantage of this type of approach is its explicit structure, which makes the system easier to implement, maintain and understand. Another advantage is that it is more efficient, as the amount of information needing to be processed decreases significantly with level. Despite these advantages, current problems include the lacks of guideline for dividing information levels and the lack of a guarantee that a globally optimal result can be achieved.

Integrated methods, on the other hand, export outputs of different conceptual levels simultaneously without partitioning the processing procedure into discrete levels. An example is a motion estimation system that deals with the whole sequence of images as a 3-D stack with two spatial coordinates and one temporal dimension [2]. This system estimates velocity without extracting feature components such as corners and boundaries. Hence a possible advantage of integrated systems is that it is unnecessary to maintain a large set of different algorithms. Also, if a good searching algorithm is integrated into this system, it is possible to force the result to a "good" local optimum or even the global optimum. In practice, the main disadvantages of this approach include the heavy computational burden and lack of understanding of the mechanism of the system.

Currently, the answer to the question of which type of structure to adopt is still an empirical one. In the next section, we compare the perfomance of these two approaches on a 1-D motion estimation problem.

Based on the results of the comparison, we develop the integrated model-based algorithm for object recognition, which can be applied to vehicle recognition in a traffic surveillance system.

2 Comparison of Hierachical and Integrated Algorithms

2.1 Construction of a 1-D Motion Estimation Model

We consider a 1-D motion estimation system in order to compare the performance of hierachical and integrated methods quantitatively. This model comprises a 1-D black background with a 1-D white object moving at a constant velocity against the background. The size of the background is L pixels, and the length of the object is l (l unknown). Given N frames of "snapshots" taken at times t_0, t_1,..., t_{N-1}, we wish to estimate the motion velocity v and the object size l. We assume the data are noise corrupted; i.e., a black pixel has probability p to be presented as a white one, and a white pixel has probability p to be presented as a black one, with all pixels independently corrupted. We also assume $L>>l$ and that the object is far from the ends of the background to avoid boundary effects. The analysis of the effects of the corruption probability p is performed for small values of p ($p < 10\%$). In the analysis we use

[1]This material is based upon work partially supported by the National Science Foundation ECS-9873451 under subontract No. SBC-MIT-5710000330, and by EPRI and the U.S. Army Research Office USARO EPRI RP 8030-15 under Subcontract to Cornell University Contract Nos. WO8333-04 and Army OSP 35352-6086. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the above agencies.

1's to represent white pixels and 0's for black pixels. Hence each "snapshot" (frame) is a vector composed of 1's and 0's, and the stack of frames is a matrix of 1's and 0's.

2.2 Hierarchical Approach

The hierarchical approach consists of three steps:
1. Identifying the object in each frame;
2. Determining the center of the object in each frame;
3. Performing linear curve fitting of the position of the center of the object with respect to time; thus, its slope is the desired velocity.

The maximum-likelihood (ML) method is adopted for identifying the white object against the black background. For a frame at time t, let the observation of pixel x be $I(x,t) \in \{0,1\}$. Assume the true endpoints of the object are located at $x_{o,t}$ and $x_{e,t}$. Hats are used to denote estimates, i.e., estimates of $x_{o,t}$ and $x_{e,t}$ are denoted by $\widehat{x_{o,t}}$ and $\widehat{x_{e,t}}$. Note that $l = x_{e,t} - x_{o,t} + 1$ and $\widehat{l} = \widehat{x_{e,t}} - \widehat{x_{o,t}} + 1$. The maximum likelihood method is:
since $1 - p > p$,

$$(\widehat{x_{o,t}}, \widehat{x_{e,t}}) = \arg \max_{x_{o,t}, x_{e,t}} \left[\left(\sum_{x_{o,t} \le x \le x_{e,t}} 2I(x,t) \right) - (x_{e,t} - x_{o,t} + 1) \right]. \quad (1)$$

The center of the object is then calculated as

$$\widehat{x_{c,t}} = \frac{(\widehat{x_{o,t}} + \widehat{x_{e,t}})}{2}. \quad (2)$$

Using the estimates for the positions of the centers for $t = 0, \ldots, N-1$, the velocity is then estimated as

$$\widehat{v} = \frac{6 \left[2 \sum_{1 \le i \le N} \left((i-1) \widehat{x_{c,i-1}} \right) - N \left(\sum_{1 \le i \le N} \widehat{x_{c,i-1}} \right) \right]}{N(N-1)(N+1)}. \quad (3)$$

2.3 Integrated Approach

If we list the frames as a time sequence, then we get a matrix composed of 0's and 1's. For the noise-free case, the 1's (white pixels) form a parallelogram under the assumption of constant velocity. The integrated method is based on this fact.

The maximum-likelihood approach is again used to estimate x_o (the position of the left end point of the object at time 0), l, and v. By a similar calculation, we obtain

$$(\overline{x_o}, \overline{l}, \overline{v}) = \arg \max_{x_o, l, v} \left(\sum_{\substack{\lfloor x_o + tv \rfloor \le x \le \lfloor x_o + (l-1) + tv \rfloor \\ 0 \le t \le N-1}} [2I(x,t) - 1] \right), \quad (4)$$

where $\lfloor \alpha \rfloor$ denotes the largest integer less than or equal to a real number α.

2.4 Analysis of the Hierarchical Algorithm

We use the magnitude of $E\left[(v - \widehat{v})^2\right]$ as the measure of performance. From (3), we have

$$E\left[(v - \widehat{v})^2\right] = \frac{36}{(N-1)^2 N^2 (N+1)^2} (A + B), \quad (5)$$

where

$$A := \sum_{1 \le i \le N} (2i - 2 - N)^2 E\left[(x_{c,t_i} - \widehat{x_{c,t_i}})^2 \right],$$

and

$$B := 2 \sum_{1 \le i < j \le N} (2i - 2 - N)(2j - 2 - N) \cdot E\left[(x_{c,t_i} - \widehat{x_{c,t_i}}) \right] E\left[\left(x_{c,t_j} - \widehat{x_{c,t_j}} \right) \right].$$

From the procedure for the ML-estimation of $\widehat{x_{c,t_i}}$ given in (1), we know that if $\widehat{x_{o,t_i}}$ and $\widehat{x_{e,t_i}}$ are selected, then we must have

$$\left(\sum_{\widehat{x_{o,t_i}} \le x \le \widehat{x_{e,t_i}}} 2I(x, t_i) \right) - (\widehat{x_{e,t_i}} - \widehat{x_{o,t_i}} + 1) \ge \left(\sum_{\widehat{x_{o,t_i}} \le x \le \widehat{x_{e,t_i}} + 1} 2I(x, t_i) \right) - ((\widehat{x_{e,t_i}} + 1) - \widehat{x_{o,t_i}} + 1),$$

or $2I(\widehat{x_{e,t_i}} + 1, t_i) \le 1$. The latter implies that $I(\widehat{x_{e,t_i}} + 1, t_i) = 0$. By similar arguments, we have the following conditions:

1. $I(\widehat{x_{e,t_i}} + 1, t_i) = I(\widehat{x_{e,t_i}} + 2, t_i) = I(\widehat{x_{o,t_i}} - 1, t_i) = I(\widehat{x_{o,t_i}} - 2, t_i) = 0$;
2. $I(\widehat{x_{e,t_i}}, t_i) = I(\widehat{x_{o,t_i}}, t_i) = 1$;
3. $I(\widehat{x_{o,t_i}} + 1, t_i) + I(\widehat{x_{o,t_i}} + 2, t_i) \ge 1$;
4. $I(\widehat{x_{e,t_i}} - 1, t_i) + I(\widehat{x_{e,t_i}} - 2, t_i) \ge 1$.

The above rules provide a means to obtain the upper bounds on the probabilities $P(\widehat{x_{o,t_i}} | x_{o,t_i})$ and $P(\widehat{x_{e,t_i}} | x_{e,t_i})$:

1. $P(\widehat{x_{o,t_i}} = x_{o,t_i}) \le (1-p)^3 (1-p^2)$
2. $P(\widehat{x_{o,t_i}} = x_{o,t_i} + 1) \le (1-p)^3 (1+p)$
3. $P(\widehat{x_{o,t_i}} = x_{o,t_i} - 1) \le (1-p)^3 (1+p)\, p$
4. $P(\widehat{x_{o,t_i}} = x_{o,t_i} - 2) \le p(1-p)^2 (1-p+p^2)$
5. $P(\widehat{x_{e,t_i}} = x_{e,t_i}) \le (1-p)^3 (1-p^2)$
6. $P(\widehat{x_{e,t_i}} = x_{e,t_i} - 1) \le (1-p)^3 (1+p)\, p$
7. $P(\widehat{x_{e,t_i}} = x_{e,t_i} + 1) \le (1-p)^3 (1+p)\, p$
8. $P(\widehat{x_{e,t_i}} = x_{e,t_i} + 2) \le p(1-p)^2 (1-p+p^2)$

The remaining situations are bounded by $O\left(p^2\right)$. Hence, we obtain the upper bounds for the probabilities of the different estmiates for x_{c,t_i} in below:

$$9.P\left(\widehat{x_{c,t_i}} = x_{c,t_i} \pm \frac{1}{2}\right) \leq 2\left(1-p\right)^7\left(1+p\right)^2 p + \left(1-p\right)^4\left(1+p\right)^2\left(1-p+p^2\right)p^2 + O\left(p^3\right)$$

$$10.P\left(\widehat{x_{c,t_i}} = x_{c,t_i} \pm 1\right) \leq \left(1-p\right)^6\left(1-p^3\right)p + 2\left(1-p\right)^6\left(1+p\right)^2 p^2 + O\left(p^3\right)$$

$$11.P\left(\widehat{x_{c,t_i}} = x_{c,t_i} \pm \frac{3}{2}\right) \leq 2\left(1-p\right)^5\left(1+p\right)^2\left(2-p\right)p^2 + O\left(p^3\right).$$

For $er > \frac{3}{2}$, $P\left(|\widehat{x_{c,t_i}} - x_{c,t_i}| = er\right)$ is dominated by $O\left(p^3\right)$. So $E\left[\left(x_{c,t_i} - \widehat{x_{c,t_i}}\right)^2\right]$ is dominated by

$$2\left(\frac{1}{2}\right)^2 2\left(1-p\right)^7\left(1+p\right)^2 p + 2\left(1-p\right)^6\left(1-p^3\right)p + O\left(p^2\right) = p\left(1-p\right)^7\left(3+4p+p^2\right) + O\left(p^2\right) \quad (6)$$

Hence A is in the order of $O\left(3p\right)$.

On the other hand, $|E\left[\widehat{x_{c,i}} - x_{c,i}\right]|$ is dominated by

$$\left|P\left(\widehat{x_{c,i}} = x_{c,i} + \frac{1}{2}\right)\frac{1}{2} + P\left(\widehat{x_{c,i}} = x_{c,i} - \frac{1}{2}\right)\left(-\frac{1}{2}\right) + P\left(\widehat{x_{c,i}} = x_{c,i}+1\right)1 + P\left(\widehat{x_{c,i}} = x_{c,i} - \frac{1}{2}\right)(-1)\right| < 2p\left(\frac{1}{2}\right) + p = 2p.$$

So $|E\left[\widehat{x_{c,i}} - x_{c,i}\right]|$ is dominated by $2p$. Thus B is dominated by a quantity of the order of $O\left(p^2\right)$.

In summary, $E\left[\left(v-\widehat{v}\right)^2\right]$ is dominated by $O\left(pN^{-3}\right)$.

The above results are verified by simulation results. As shown in Figure 1(a), $\ln E\left[\left(v-\widehat{v}\right)^2\right]$ versus $\ln N$ is a straight line with slope close to -3, as predicted. Figure 1(b) shows the relationship between $E\left[\left(v-\widehat{v}\right)^2\right]$ and p. We see that the slope of the line $\ln E\left[\left(v-\widehat{v}\right)^2\right]$ versus $\ln p$ is close to 1 for p small ($< 10\%$). When p gets larger, the performance is worse than that specified in (5).

2.5 Analysis of the Integrated Method

First we discuss some issues affecting the error due to integer rounding. A careful study shows that if we list all the rational numbers in the form $\frac{m}{n}$ (m, n are integers with $|m| \leq n$ and $1 \leq n \leq N-1$) in ascending order to get a sequence $-1 = a_0 < a_1 < \ldots < a_{M-2} < a_{M-1} = 1$, then the integrated algorithm cannot differentiate different velocities within the interval $[K+a_i, K+a_{i+1})$ for any integer K and $i \in \{0, 1, \ldots, M-1\}$ (set $a_M = a_0 + 2$). Hence, in the following analysis, we only consider the case where $v_{true} = K + a_i$ for some $i \in \{0, 1, ..., M-1\}$. Also, $E\left[\left(v-\widehat{v}\right)^2\right]$ could range from $O\left(N^{-2}\right)$ (e.g., for $v_{true} = \frac{1}{N}$) to $O\left(N^{-4}\right)$ (e.g., for $v_{true} = \frac{3N-5}{2(N-2)(N-1)}$), depending on the true velocity.

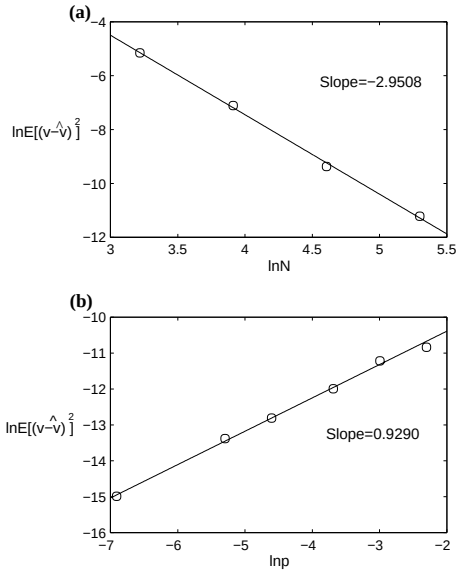

Figure 1: The simulation results for the hierarchical method. (a) The plot of $\ln E\left[\left(v-\widehat{v}\right)^2\right]$ vs. $\ln N$. Simulation is performed under the condition $L = 200$, $l = 20$, $v_{true} = 0$, $p = 0.05$. (b) The plot of $\ln E\left[\left(v-\widehat{v}\right)^2\right]$ vs. $\ln p$. Simulation is performed under the condition $L = 200$, $l = 20$, $v_{true} = 0$, $N = 200$. Both simulations are repeated 200 times for statistical purposes.

Next we show that in the estimation process, it is not likely that $\widehat{x_o} \neq x_o$ or $\widehat{x_e} \neq x_e$ (here x_o and x_e are the positions of the end points of the object at time $t = 0$). Note that for $\widehat{v} = v$, $\widehat{x_e} = x_e + 1$, and $\widehat{x_o} = x_o$, we will need to have $2\sum_{0 \leq i \leq N-1} I(x_e + iv + 1, i-1) - N > 0$, or $\sum_{0 \leq i \leq N-1} I(x_e + iv + 1, i-1) > \frac{N}{2}$. The probability of this is $O\left(p^{\frac{N}{2}}\right)$, which can be ignored for large N. Similar results can be obtained for the other situations when $\widehat{x_o} \neq x_o$ or $\widehat{x_e} \neq x_e$. Thus in the following analysis we can safely assume $\widehat{x_o} = x_o$ and $\widehat{x_e} = x_e$.

Now we consider the probabilities of each possible error. If $v_{true} = K + a_i$, then the most likely false estimates are $\widehat{v} = K + a_{i-1}$ and $\widehat{v} = K + a_{i+1}$. For each of the two cases, there is an integer $k \in \{0, 1, \ldots N-1\}$ such that for the kth frame we have $\sum_{0 \leq j \leq N-1} I(x_i - 1 + kv +$

$j,k) > \sum_{0\le j\le N-1} I(x_i+kv+j,k)$ or $\sum_{0\le j\le N-1} I(x_i+1+kv+j,k) > \sum_{0\le j\le N-1} I(x_i+kv+j,k)$. This means that either $I(x_i-1+kv,k) > I(x_e+kv,k)$ or $I(x_e+1+kv,k) > I(x_i+kv,k)$. The former implies $I(x_{i-1}+kv,k)=1$ and $I(x_e+kv,k)=0$; the latter implies $I(x_{e+1}+kv,k)=1$ and $I(x_i+kv,k)=0$. For each case, the probability is p^2. Similar analysis shows that the probability that $\widehat{v} = K + a_j$ (for $|j-i|>1$) is of the order of $O\left(p^4\right)$. So $E\left[(v-\widehat{v})^2\right]$ is dominated by $(a_i-a_{i-1})^2 p^2 + (a_i-a_{i+1})^2 p^2$, which is $O\left(p^2\right)$.

2.6 Remarks

In the above sections, both hierachical and integrated methods display their own merits. The integrated method shows a higher accuracy as $E\left[(v-\widehat{v})^2\right]$ for the integrated method is $O\left(p^2\right)$, while $E\left[(v-\widehat{v})^2\right]$ for the hierachical approach is $O(p)$. On the other hand, the hierarchical method also has its advantages. First, it requires less prior knowledge. In this algorithm we have not used the assumption that the length of the object is fixed. Second, this method is more flexible since it does not require the "constant velocity" condition until the last step for calculating the object velocity. Third, the computational complexity of the hierarchical method is lower. The number of sets of parameters needing to be tested in each frame in the hierarchical method is $\frac{(N-1)N}{2}$, so the total number of parameter sets is $O\left(N^3\right)$. For the integrated method, the number of possible $(\widehat{x_o},\widehat{x_e})$ pairs is $\frac{(N-1)N}{2}$ and the possible number of values of $\widehat{v}$ is $O(M)$ (M is as defined in Section 2.5, so $M \propto N^2$). Hence the total number of parameter sets is $O\left(N^4\right)$.

As a hierarchical method has an advantage in tracking moving object under complex conditions (e.g., varying velocity, turning road), it can be used in the vehicle tracking system. On the other hand, for processes such as model-matching, we will use the integrated algorithm for high accuracy.

3 Object Recognition Using an Integrated Approach

Because vehicle classification is one goal of our system, model matching is inevitable at some stage of processing. We adopt a 3-D deformable block model instead of the widely used wire-frame model to avoid the influence of illumination, noise and other factors. We assume that most of the vehicles have a different light density than the background. This deformable model (a parametric, generic model) is for general vehicles. After the match, we obtain information about the size, location, and orientation of the vehicle. Hence this matching process is indeed an integrated approach.

3.1 Optimization by Using Simulated Annealing

As described above, we are in fact dealing with an optimization problem having a large number of variables. In order to find the best match, we use "simulated annealing"(SA) to deal with the function which has a large number of variables and multiple local minima. A cooling schedule that ensures the SA algorithm converges to a global minimal point in probability is chosen according to [3, 4, 5, 6].

3.2 Matching of Simple Models

We first try to identify a white rectangle against a black background. The data (pixels in the image) are corrupted by noise in the same manner as in Section 2, i.e., $P\left(I(x,y)=0|(x,y)_{true}=1\right) = P\left(I(x,y)=1|(x,y)_{true}=0\right) = p < \frac{1}{2}$. An example of such data is shown in Figure 2. We have five parameters to estimate: the dimensions of the rectangle $\left(\overline{a},\overline{b}\right)$ with $\overline{a} \ge \overline{b}$, the position of the center of the object $(\overline{x},\overline{y})$, and the orientation θ, which is the angle between the longitudal axis of the object and the x-axis of the image. Due to symmetry of the object, we take $\theta \in [0,\pi)$. By a maximum-likelihood analysis similar to that in Section 2.2, we find

$$\left(\widehat{a},\widehat{b},\widehat{x},\widehat{y},\widehat{\theta}\right) = \arg\max f\left(a,b,x,y,\theta\right)$$

where $\left(\widehat{a},\widehat{b},\widehat{x},\widehat{y},\widehat{\theta}\right)$ are the estimates for $\left(\overline{a},\overline{b},\overline{x},\overline{y},\overline{\theta}\right)$, and

$$f\left(a,b,x,y,\theta\right) = ab - 2 \sum_{(u,v)\ inside\ rectangle\ defined\ by\ (a,b,x,y,\theta)} I(u,v).$$

The initial estimates x_0 and y_0 for $\overline{x}$ and $\overline{y}$ are obtained by studying $\sum_{1\le i\le n} I(i,y)$ versus y and $\sum_{1\le j\le n} I(x,j)$ versus x. We expect that the columns containing the object pixels have larger sum of pixel values than those columns forming the background. Thus, by appropriate thresholding, we pick the midpoint of the longest segment with largest $\sum_{1\le i\le n} I(i,y)$ value as x_0, and the length of the segment as an estimate for the size of one dimension (e.g., a_0). Similarly we can use the $\sum_{1\le j\le n} I(x,j)$ versus x graph to obtain an estimate y_0 and the size of another dimension (e.g., b_0). We set $\theta_0 = \frac{\pi}{2}$.

With this as the initial condition, the search is then performed by simulated annealing and is terminated after a certain fixed number of iterations (e.g., 2000). Figure 2 displays an example of the experimental data.

3.3 Extension to a 3-D Model

We next consider objects that are cubes with three dimensions $a \ge b \ge c$. The location of the center is (x_0,y_0,z_0) and its orientation is determined

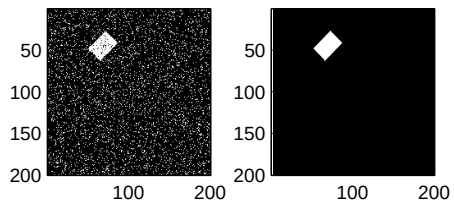

Figure 2: Matching results for a rectangular object. The left panel is the noisy image, and the right panel is the detected rectangle. The noise is 10% and $a = 30, b = 20, \theta = 45^o$.

by (α, β, γ), where α, β, and γ are the angles between the longitudal axis of the object and the x, y, and z-axes, respectively. The image is assumed to be the orthographic projection of this object onto the $x - y$ image plane. We would like to estimate $(a, b, c, x_0, y_0, z_0, \alpha, \beta, \gamma)$. Due to the nature of orthographic projection, we cannot estimate z_0. So we just set $z = 0$ for convenience. The pixels of the projection are white and the remaining pixels are black. The rest of the problem is the same as in Section 3.1.

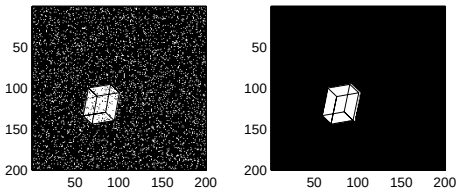

Figure 3: Result for matching a cube with $a = 30$, $b = 20$, and $c = 10$. The left panel is the synthesized image. The right panel is the matching result.

The result of applying our algorithm is shown in Figure 3. It reflects satisfactory performance of the algorithm.

3.4 Construction of a 3-D Deformable Vehicle Model

Deformable wire-frame models have been adopted in several vehicle tracking systems [7, 8] due to their flexibility. Next we develop our own deformable block model that uses dimensional information as parameters.

Our deformable vehicle model comprises three components: front, center, and rear. The front part corresponds to the hood area of the car or the extruded engine containment of the trucks and some school buses. The center is that part of the vehicle that has all the windows; it corresponds to the main part of vans, cars, buses, and the driver's cab for trucks. The rear part corresponds to the trunk of cars or the trailer of trucks. Figure 4 shows the three parts and the related notations.

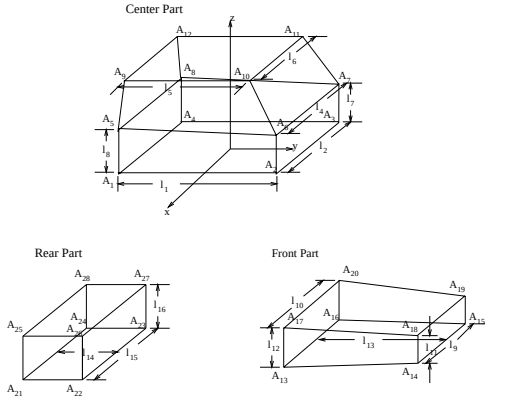

Figure 4: The three parts of the deformable vehicle model.

The parametrization of the model is based on some essential assumptions derived from our knowledge and common sense. The first assumption is that all vehicles are left-right symmetric with respect to a plane perpendicular to the ground plane. The second assumption is the ground plane constraint which says the vertices on the bottom of the vehicle (A_1, A_2, A_3, A_4, A_{13}, A_{14}, A_{15}, A_{16}, A_{21}, A_{22}, A_{23}, and A_{24}) are in the same plane that is parallel to the ground plane. The third assumption is that several planes in this vehicle model can be approximated by rectangles. The planes $A_1A_2A_3A_4$, $A_{21}A_{22}A_{23}A_{24}$, $A_9A_{10}A_{11}A_{12}$, and $A_{25}A_{26}A_{27}A_{28}$ are considered rectangles parallel to the ground plane; planes $A_{22}A_{23}A_{27}A_{26}$, $A_{21}A_{22}A_{26}A_{25}$, $A_{24}A_{21}A_{25}A_{28}$, $A_{23}A_{24}A_{28}A_{27}$, $A_2A_3A_7A_6$, $A_1A_4A_8A_5$, $A_{16}A_{13}A_{17}A_{20}$, and $A_{14}A_{15}A_{19}A_{18}$ are rectangles perpendicular to the ground plane. The plane $A_5A_6A_7A_8$ is a rectangle with variable orientation. Besides the assumptions, additional constraints are listed in Table 1.

Now we can define the vehicle coordinate system. The origin is chosen to be at the center of rectangle $A_1A_2A_3A_4$. The x, y, and z directions correspond to the rightward, forward, and upward directions, respectively. The center of $A_9A_{10}A_{11}A_{12}$ is at (x', y', h). From the symmetry assumption we know that $x' = 0$. Hence the coordinates of each of the vertices can be determined.

By putting together all the constraints and descriptions, we obtain the deformable model of the general vehicle. This model can be used to generate different types of vehicles by varying the parameteres.

3.5 Matching the Deformable Model with the Image Data

To match the deformable model, we need to tune 15 independent internal parameters (l_1, l_2, l_5, l_6, l_7, l_8, l_9, l_{10}, l_{11}, l_{13}, l_{14}, l_{15}, l_{16}, h, and y') and several external parameters. Under the ground plane constraint, the number of external parameters is reduced to 3 (the

Table 1: Geometrical constraints in the vehicle model

Constraint	Meaning
$l_7 = l_{12} \geq l_{11}$	The hood is attached to the bottom of the windshield
$1.4l_8 \leq h \leq 1.8l_8$	The height of the windows is in the right range
$l_{16} \geq l_8$	The trunk (trailer) is not lower than the bottom of rear window
$1.5l_2 \geq l_{15} \geq l_2$	The trunk is not narrower than the body and not too wide
$l_2 \geq l_{10} \geq l_9$	The hood is not wider than the body
$l_2 = l_4$	Geometric constraints
$l_6 \leq l_2$, $l_5 \leq l_1$	The size of the top of the vehicle is no larger than the bottom of the body
$y' + \frac{l_5}{2} \leq \frac{l_1}{2}$	The front window tilts backward
$-y' + \frac{l_5}{2} \leq \frac{l_1}{2}$	The rear window tilts forward

location (x_0, y_0) of the origin of the vehicle frame and θ-the angle between the y-axis of the vehicle frame and the y-axis of the ground frame). Thus the total number of parameters is 18.

The result of matching the deformable model with a synthesized images are shown in Figure 5. The data are images with the same settings as in previous section. We assume that the external camera calibration information is available. The most significant feature of

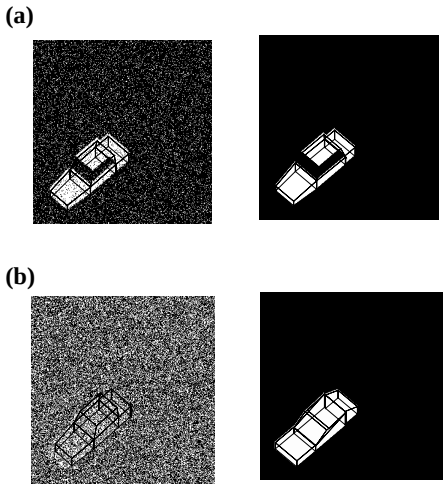

Figure 5: An example of matching the deformable model with the synthetic image. (a) The noise is 10%. (b)The noise is 40%.

this algorithm is its tolerance to noise. In Figure 5(b), although the restored model is deformed in the presence of $p = 40\%$ noise, its general contour still fits well with the original model and can be easily recognized.

4 Summary

We have constructed a deformable model for vehicles and explored the possibility of exploiting this model to fit the image data. Most of the results are positive. However, we do observe some problems with matching. In practice, when matching the model to a bright-colored bus with black window, the model does not converge to the right position (data not shown). Analysis shows that we need to modify our cost function to avoid it. The long running time is another problem. Possible remedies include adding more constraints based on prior knowledge to reduce the search space, applying geometrical constraints for multiple view matching, and simplifying the optimization algorithm.

Our scheme for matching fits well with the notion of the integrated approach in Section 2. The outputs such as sizes and orientation can be used for purposes such as classification, velocity estimation and tracking. Hence this algorithm can be adopted for the development of a traffic surveillance system.

References

[1] Granlund, G.H., "The complexity of vision," *Signal Processing,* Vol. 74, pp101-126, 1999.

[2] Jahne, B., "Motion determination in space-time images," in *SPIE-Image Processing III,* Vol.1135, 1989, pp147-152.

[3] Mitra, D., Romeo, F., and Sangiovanni-Vincentelli, A., "Convergence and finite-time behavior of simulated annealing." In *Proceedings of the IEEE Conference on Decision and Control Including The Symposium on Adaptive Processes.* , 1986, pp 761-767.

[4] Hajek, B., "Cooling schedules for optimal annealing." *Mathematics of Operations Research,* vol. 13, no. 2, pp311-329, 1988.

[5] Tsitsiklis, J.M., "Markov chains with rare transitions and simulated annealing,"*Preprint, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology,* August 1985.

[6] Connors, D.P., and Kumar, P.R., "Balance of recurrence order in time-inhomogeneous markov chains with application to simulated annealing," *Probability in the Engineering and Informational Sciences* no. 2, pp157-184, 1988.

[7] Ferryman, J.M., Worrall, A.D., Sullivan, G.D., and Baker, K.D., "Visual surveillance using deformable models of vehicles," *Elsevier. Robotics and Autonomous Systems,* vol.19, no.3-4, pp.315-335, 1997.

[8] Dubuisson, M.-P., and Jain, A.K., "Contour extraction of moving objects in complex outdoor scenes," *International Journal of Comptuer Vision,* vol.14, pp83-105, 1995.